



AF
SW

TRANSMITTAL OF APPEAL BRIEF			Docket No. BBNT-P01-199
In re Application of: Colbath et al.			
Application No. 10/685,479	Filing Date October 16, 2003	Examiner M. M. Pyo	Group Art Unit 2161
Invention: SYSTEMS AND METHODS FOR LABELING CLUSTERS OF DOCUMENTS			

TO THE COMMISSIONER OF PATENTS:

Transmitted herewith is the Appeal Brief in this application, with respect to the Notice of Appeal filed: June 5, 2007.

The fee for filing this Appeal Brief is \$ 500.00.

☒ Large Entity ☐ Small Entity

☐ A petition for extension of time is also enclosed.


The fee for the extension of time is _____.

☐ A check in the amount of _____ is enclosed.

☒ Charge the amount of the fee to Deposit Account No. 18-1945.
This sheet is submitted in duplicate.

☐ Payment by credit card. Form PTO-2038 is attached.

☒ The Director is hereby authorized to charge any additional fees that may be required or credit any overpayment to Deposit Account No. 18-1945.
This sheet is submitted in duplicate.



Edward A. Gordon
Attorney Reg. No. : 54,130
FISH & NEAVE IP GROUP, ROPES & GRAY LLP
1211 Avenue of the Americas
New York, New York 10036
(617) 951-7066

Dated: August 6, 2007

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being deposited with the U.S. Postal Service on the date shown below with sufficient postage as First Class Mail, in an envelope addressed to: MS Appeal Brief - Patents, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

Dated: 8/16/07 Signature: Joanne Ryan (Joanne Ryan)



I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being deposited with the U.S. Postal Service on the date shown below with sufficient postage as First Class Mail, in an envelope addressed to:
MS Appeal Brief - Patents, Commissioner for Patents, P.O. Box 1450,
Alexandria, VA 22313-1450.

Dated: 8/6/07

Signature: Joanne Ryan

(Joanne Ryan)

Docket No.: BBNT-P01-199
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:
Colbath et al.

Application No.: 10/685,479

Confirmation No.: 5448

Filed: October 16, 2003

Art Unit: 2161

For: SYSTEMS AND METHODS FOR LABELING
CLUSTERS OF DOCUMENTS

Examiner: M. M. Pyo

APPEAL BRIEF

MS Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

This Appeal Brief is submitted in response to the Final rejection mailed February 5, 2007 and in support of the Notice of Appeal filed June 5, 2007.

This brief contains items under the following headings as required by 37 C.F.R. § 41.37 and M.P.E.P. § 1205.2:

- | | |
|-------|---|
| I. | Real Party In Interest |
| II | Related Appeals and Interferences |
| III. | Status of Claims |
| IV. | Status of Amendments |
| V. | Summary of Claimed Subject Matter |
| VI. | Grounds of Rejection to be Reviewed on Appeal |
| VII. | Arguments |
| VIII. | Conclusion |
| IX. | Claims |
| X. | Evidence |
| XI. | Related Proceedings Appendix |

I. REAL PARTY IN INTEREST

The real party in interest for this appeal is BBN Technologies Corp.

II. RELATED APPEALS AND INTERFERENCES

Appellants are unaware of any related appeals, interferences or judicial proceedings.

III. STATUS OF CLAIMS

Claims 1-19 are pending in this application. Claims 12-15 have been withdrawn from consideration and claims 1-11 and 16-19 have been rejected. Claims 1-11 and 16-19 are the subject of the present appeal.

IV. STATUS OF AMENDMENTS

No Amendment has been filed subsequent to the Final Office Action mailed February 5, 2007.

V. SUMMARY OF CLAIMED SUBJECT MATTER

Each of the independent claims involved in this appeal is recited below, followed in parenthesis by examples of where support can be found in the specification and drawings for the claimed subject matter. In addition, each dependent claim argued separately below is also summarized in a similar manner.

Claim 1 recites: A method of creating labels for clusters of documents, comprising: identifying topics associated with the documents in the clusters (e.g., 710; Fig. 7, pg. 12, lines 18-20); determining whether the topics are associated with at least half of the documents in the clusters (e.g., 720, Fig. 7; pg. 12, lines 5-7); adding ones of the topics that are associated with at least half of

the documents in the clusters to cluster lists (e.g., 730, Fig. 7; pg. 13, lines 7-8); and forming labels for the clusters from the cluster lists (e.g., 750, Fig. 7; pg. 13, lines 12-13).

Claim 4 recites: The method of claim 3, wherein the ranking the ones of the topics includes: assigning ranks to the ones of the topics based on a number of the documents with which the ones of the topics are associated (e.g., pg. 13, lines 13-15).

Claim 5 recites: The method of claim 1, further comprising: ranking the ones of the topics based on a number of the documents with which the ones of the topics are associated (e.g., pg. 13, lines 12-15).

Claim 9 recites: A system for generating a label for a cluster of documents, comprising: means for identifying topics associated with the documents in the cluster (e.g., 520, Fig. 5; pg. 11, lines 11-12; pg. 12, lines 18-20); means for determining whether the topics are associated with at least half of the documents in the cluster (e.g., 520, Fig. 5; pg. 11, lines 12-13; pg. 13, lines 5-7); and means for generating a label for the cluster based on one or more of the topics that are associated with at least half of the documents in the cluster (e.g., 520, Fig. 5; pg. 11, lines 12-13; pg. 13, lines 7-8 and 12-13).

Claim 16 recites: A topic detection system, comprising: a decision engine (e.g., 510, Fig. 5) configured to: receive a plurality of documents (e.g., pg. 10, line 18), and group the documents into a plurality of clusters (e.g., pg. 10, line 21 – pg. 11, line 3); and a label engine (e.g. 520, Fig. 5)

configured to: identify topics associated with the documents in the clusters (e.g., pg. 11, lines 11-12), determine whether the topics are associated with at least half of the documents in the clusters (e.g., pg. 11, lines 12-13), and form labels for the clusters using ones of the topics that are associated with at least half of the documents in the clusters (e.g., pg. 11, lines 12-13).

Claim 17 recites: The system of claim 16, wherein the label engine is further configured to: rank the one of topics based on a number of the documents with which the ones of the topics are associated (e.g., pg. 13, lines 12-15).

Claim 18 recites: A method for creating labels for clusters of documents, comprising: identifying topics associated with the documents in the clusters (e.g., 710, Fig. 7; pg. 12, lines 18-20); determining whether the topics are associated with at least a predetermined number of the documents in the clusters (e.g., 720, Fig. 7; pg. 13, lines 5-7); and generating labels for the clusters using ones of the topics that are associated with the at least a predetermined number of the documents in the clusters (e.g., 730, 750; Fig. 7; pg. 12, lines 7-8, 12-13).

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

Claims 1-3, 9, and 16 have been rejected under 35 U.S.C. § 102(b) as being anticipated by Colbath et al. ("Spoken Documents: Creating Searchable Archives from Continuous Audio," 2000) (which the Examiner identified as "Kubala et al."); claims 4-8, 10, 11, and 17 have been rejected under 35 U.S.C. § 103(a) as unpatentable over Colbath et al. in view of Liddy et al. (U.S. Patent No. 5,963,940); and claims 18 and 19 have been rejected under 35 U.S.C. § 102(b) as anticipated by, or in the alternative, under U.S.C. § 103(a) as unpatentable over Colbath et al.

VII. ARGUMENT

A. Rejection under 35 U.S.C. § 102 based on Colbath et al.

The initial burden of establishing a *prima facie* basis to deny patentability to a claimed invention always rests upon the Examiner. In re Oetiker, 977 F.2d 1443, 24 U.S.P.Q.2d 1443 (Fed. Cir. 1992). A proper rejection under 35 U.S.C. § 102 requires that a single reference teach every aspect of the claimed invention. Any feature not directly taught must be inherently present. Verdegaal Bros. v. Union Oil Co. of California, 814 F.2d 628, 2 USPQ2d 1051 (Fed. Cir. 1987).

1. Claims 1-3

Independent claim 1 recites a method of creating labels for clusters of documents. The method includes identifying topics associated with the documents in the clusters; determining whether the topics are associated with at least half of the documents in the clusters; adding ones of the topics that are associated with at least half of the documents in the clusters to cluster lists; and forming labels for the clusters from the cluster lists.

Colbath et al. does not disclose or suggest the combination of features recited in claim 1. For example, Colbath et al. does not disclose clusters of documents. Colbath et al. mentions the word "cluster" in a few places in the context of clustering speakers together (see, e.g., column 5, lines 21-32), but does not disclose clusters of documents.

Because Colbath et al. does not disclose clusters of documents, Colbath et al. cannot disclose determining whether topics are associated with at least half of the documents in the clusters, as further recited in claim 1. The Examiner alleged that Colbath et al. discloses this feature and cited

column 14, lines 1-26, of Colbath et al. for support (final Office Action, page 3). Appellants disagree.

At column 14, lines 1-26, Colbath et al. discloses:

One solution to this is to use traditional relevance feedback. The user has the option of specifying an entire story to the query system. When this is done, all the words in the story are fed into the full-text search engine, which returns five documents that use the maximum number of common terms with the seed document.

In the example in Fig. 7, we've given the system the first story in the "Smoking and FDA" query for a relevance feedback operation. The full-text search system has returned five stories. The first one is the seed story, since it has the most terms in common with itself. The second one happens to be the second-ranked story from the boolean query. The remaining three stories, however, are three stories on highly similar topics that weren't found with the boolean query mechanism. It should be emphasized that this mode of search becomes particularly important when the document source is text with errors introduced by a speech recognition system. Because of speech recognition errors, highly relevant documents may fall through the cracks of a boolean search, but are more likely to be found via relevance feedback since they will contain other words in common that are recognized correctly.

In this section, Colbath et al. discloses the use of traditional relevance feedback involving feeding all the words in a story into a full-text search engine, which returns five documents that use the maximum number of common terms with the seed document. Nowhere in this section, or elsewhere, does Colbath et al. disclose or suggest determining whether identified topics are associated with at least half of the documents in the clusters, as required by claim 1.

The Examiner also alleged that "three out of five search results contain highly similar topics for the cluster group" (final Office Action, page 3). Appellants cannot understand what the Examiner is alleging, but assert that, even if this allegation is reasonable, it has nothing to do with determining whether identified topics are associated with at least half of the documents in the clusters, as required by claim 1.

Colbath et al. also does not disclose forming labels for the clusters from cluster lists that include topics that are associated with at least half of the documents in the clusters, as further recited in claim 1. Instead, Colbath et al. discloses forming a title for a single document from all of the topics associated with that document (column 10, lines 13-20).

The Examiner alleged that Colbath et al. discloses forming labels for the clusters from cluster lists that include topics that are associated with at least half of the documents in the clusters, and cited column 15, lines 27-37, column 8, lines 11-16, column 14, line 29 - column 15, line 25, of Colbath et al. for support (final Office Action, page 3). Appellants disagree.

At column 15, lines 27-37, Colbath et al. discloses:

There is no particular reason that this technique could not be extended to include proper name tagging, marking of new vocabulary words for the recognizer, or identification of new topics for the topic classifier. The latter is particularly important, since it is unlikely that an end-user could find a ready-made set of topics for their own meetings or teleconferences. For some particular problem domains, it may be sufficient to have a small set of topics (3-4 instead of the current 5,500).

In this section, Colbath et al. discloses implementing techniques to include proper name tagging, marking of new vocabulary words for a recognizer, or identification of new topics for a topic classifier. Nowhere in this section, or elsewhere, does Colbath et al. disclose or suggest forming labels for the clusters from cluster lists that include topics that are associated with at least half of the documents in the clusters, as required by claim 1.

At column 8, lines 11-16, Colbath et al. discloses:

The Rough'n'Ready IR system uses a full-text search system developed at BBN which uses an HMM-based model of document retrieval. This system, described in [7], is used in relevance-feedback mode to allow the user of the system to find documents that are similar to an exemplar.

In this section, Colbath et al. discloses an HMM-based model for document retrieval. Nowhere in this section, or elsewhere, does Colbath et al. disclose forming labels for the clusters from cluster lists that include topics that are associated with at least half of the documents in the clusters, as required by claim 1.

At column 14, lines 29 - column 15, line 25, Colbath et al. discloses:

Annotation

There is no particular reason that the database has to be browsed in a read-only fashion, however. The training data for the Rough'n'Ready indexer is currently fairly static. To annotate more speech data, or additional names for the name spotter, or additional topics for the topic classifier is a separate, offline process using dedicated annotators. However, since the current annotation process is relatively simple and does not require any in-depth linguistic knowledge, it seems logical that the end user of the archive should be enlisted in helping to provide the training data. This makes sense since it is likely the consumer of the data will have the most familiar with it, and will be able to provide topics, identify speakers, etc.

The current Rough'n'Ready system includes some basic speaker annotation capabilities. If the user encounters a speaker currently marked as unknown, they can step through a relatively simple wizard that will play segments of data that have been tagged with the same identifier (such as "Male 5") and ask them to confirm that this is the same as the first speaker. Once they have accumulated enough data (three to five minutes), the system trains a new speaker model, and reprocesses the rest of the archive off-line to include the new speaker. It is also possible to add extra training data for speakers that have particularly weak performance, improving their models.

In this section, Colbath et al. discloses that annotators can annotate additional speech data, additional names for the name spotter, or additional topics for the topic classifier, and that the current Rough'n'Ready system includes some basic speaker annotation capabilities. Nowhere in this section, or elsewhere, does Colbath et al. disclose or suggest forming labels for the clusters from cluster lists that include topics that are associated with at least half of the documents in the clusters, as required by claim 1.

The Examiner also alleged that Colbath et al. discloses a "new identification of new topics for the topic classifier" (final Office Action, page 3). Even assuming, for the sake of argument, that

the Examiner's allegation is reasonable (a point that Appellants do not concede), the Examiner has not addressed the feature of forming labels for the clusters from cluster lists that include topics that are associated with at least half of the documents in the clusters, as required by claim 1.

In response to the Examiner's arguments on page 8 of the final Office Action that

it should be noted that Kubula discloses the speaker identification and segmentation system creating paragraph-like units between speakers and clustering archived files with a unique name. Thus, the Kubula's teaching of processing a clustered archived files with unique names is valid to read on the broadly claimed limitation of "identifying topics associated with the documents in the clusters,"

Appellants respectfully disagree. Speaker identification and segmentation allows the system to detect changes between speakers, which is important for correct playback of audio sections of an archive (Colbath et al., column 5, lines 19-23). Speaker identification has nothing to do with "identifying topics associated with the documents in the clusters," as recited in claim 1.

Even assuming, for the sake of argument, that speaker identification and segmentation could reasonably be equated to identifying topics associated with the documents in the clusters (a point that Appellants do not concede), nowhere does Colbath et al. disclose or suggest that the "unique name" assigned to the archived files are formed from cluster lists that include topics that are associated with at least half of the documents in the clusters, as required by claim 1.

The Examiner continues to argue that Colbath et al. discloses the recited features of claim 1, but merely restates the previous rejection without explaining how the cited sections of Colbath et al. disclose the recited features of claim 1.

For at least the foregoing reasons, Appellants submit that the rejection of claim 1 under 35 U.S.C. § 102(b) based on Colbath et al. is improper. Accordingly, Appellants request that the rejection be reversed.

Claims 2 and 3 depend from claim 1. Therefore, Appellants request that the rejection of claims 2 and 3 be reversed for at least the reasons given above with respect to claim 1.

2. Claim 9

Claim 9 recites a system for generating a label for a cluster of documents. The system includes means for identifying topics associated with the documents in the cluster; means for determining whether the topics are associated with at least half of the documents in the cluster; and means for generating a label for the cluster based on one or more of the topics that are associated with at least half of the documents in the cluster. Colbath et al. does not disclose or suggest this combination of features.

For example, Colbath et al. does not disclose clusters of documents. Colbath et al. mentions the word "cluster" in a few places in the context of clustering speakers together (see, e.g., column 5, lines 21-32), but does not disclose clusters of documents.

Because Colbath et al. does not disclose clusters of documents, Colbath et al. cannot disclose means for determining whether topics are associated with at least half of the documents in the clusters, as recited in claim 9. The Examiner alleged that Colbath et al. discloses this feature and cited column 14, lines 1-26, of Colbath et al. for support (final Office Action, page 3). Appellants disagree.

At column 14, lines 1-26, Colbath et al. discloses:

One solution to this is to use traditional relevance feedback. The user has the option of specifying an entire story to the query system. When this is done, all the words in the story are fed into the full-text search engine, which returns five documents that use the maximum number of common terms with the seed document.

In the example in Fig. 7, we've given the system the first story in the "Smoking and FDA" query for a relevance feedback operation. The full-text search system has returned five stories. The first one is the seed story, since it has the most terms in common with itself. The second one happens to be the second-ranked story from the boolean query. The remaining three stories, however, are three stories on highly similar topics that weren't found with the boolean query mechanism. It should be emphasized that this mode of search becomes particularly important when the document source is text with errors introduced by a speech recognition system. Because of speech recognition errors, highly relevant documents may fall through the cracks of a boolean search, but are more likely to be found via relevance feedback since they will contain other words in common that are recognized correctly.

In this section, Colbath et al. discloses the use of traditional relevance feedback involving feeding all the words in a story into a full-text search engine, which returns five documents that use the maximum number of common terms with the seed document. Nowhere in this section, or elsewhere, does Colbath et al. disclose or suggest means for determining whether identified topics are associated with at least half of the documents in the clusters, as required by claim 1.

The Examiner also alleged that "three out of five search results contain highly similar topics for the cluster group" (final Office Action, page 3). Appellants cannot understand what the Examiner is alleging, but assert that, even if this allegation is reasonable, it has nothing to do with determining whether identified topics are associated with at least half of the documents in the clusters, as required by claim 1.

Colbath et al. also does not disclose means for generating a label for the cluster based on one or more of the topics that are associated with at least half of the documents in the cluster, as further recited in claim 9. Instead, Colbath et al. discloses forming a title for a document from all of the topics associated with that document (column 10, lines 13-20).

The Examiner alleged that Colbath et al. discloses means for generating a label for the cluster based on one or more of the topics that are associated with at least half of the documents in

the cluster, and cited column 15, lines 27-37, column 8, lines 11-16, column 14, line 29 - column 15, line 25, of Colbath et al. for support (final Office Action, page 3). Appellants disagree.

At column 15, lines 27-37, Colbath et al. discloses:

There is no particular reason that this technique could not be extended to include proper name tagging, marking of new vocabulary words for the recognizer, or identification of new topics for the topic classifier. The latter is particularly important, since it is unlikely that an end-user could find a ready-made set of topics for their own meetings or teleconferences. For some particular problem domains, it may be sufficient to have a small set of topics (3-4 instead of the current 5,500).

In this section, Colbath et al. discloses implementing techniques to include proper name tagging, marking of new vocabulary words for a recognizer, or identification of new topics for a topic classifier. Nowhere in this section, or elsewhere, does Colbath et al. disclose or suggest means for generating a label for the cluster based on one or more of the topics that are associated with at least half of the documents in the cluster, as required by claim 9.

At column 8, lines 11-16, Colbath et al. discloses:

The Rough'n'Ready IR system uses a full-text search system developed at BBN which uses an HMM-based model of document retrieval. This system, described in [7], is used in relevance-feedback mode to allow the user of the system to find documents that are similar to an exemplar.

In this section, Colbath et al. discloses an HMM-based model for document retrieval. Nowhere in this section, or elsewhere, does Colbath et al. disclose or suggest means for generating a label for the cluster based on one or more of the topics that are associated with at least half of the documents in the cluster, as required by claim 9.

At column 14, lines 29 - column 15, line 25, Colbath et al. discloses:

Annotation

There is no particular reason that the database has to be browsed in a read-only fashion, however. The training data for the Rough'n'Ready indexer is currently fairly static. To annotate more speech data, or additional names for the name spotter, or additional topics for the topic classifier is a separate, offline process using dedicated annotators. However, since

the current annotation process is relatively simple and does not require any in-depth linguistic knowledge, it seems logical that the end user of the archive should be enlisted in helping to provide the training data. This makes sense since it is likely the consumer of the data will have the most familiar with it, and will be able to provide topics, identify speakers, etc.

The current Rough'n'Ready system includes some basic speaker annotation capabilities. If the user encounters a speaker currently marked as unknown, they can step through a relatively simple wizard that will play segments of data that have been tagged with the same identifier (such as "Male 5") and ask them to confirm that this is the same as the first speaker. Once they have accumulated enough data (three to five minutes), the system trains a new speaker model, and reprocesses the rest of the archive off-line to include the new speaker. It is also possible to add extra training data for speakers that have particularly weak performance, improving their models.

In this section, Colbath et al. discloses that annotators can annotate more speech data, additional names for the name spotter, or additional topics for the topic classifier, and that the current Rough'n'Ready system includes some basic speaker annotation capabilities. Nowhere in this section, or elsewhere, does Colbath et al. disclose or suggest means for generating a label for the cluster based on one or more of the topics that are associated with at least half of the documents in the cluster, as required by claim 9.

The Examiner also alleged that Colbath et al. discloses a "new identification of new topics for the topic classifier" (final Office Action, page 3). Even assuming, for the sake of argument, that the Examiner's allegation is reasonable (a point that Appellants do not concede), the Examiner has not addressed the feature of means for generating a label for the cluster based on one or more of the topics that are associated with at least half of the documents in the cluster, as required by claim 9.

In response to the Examiner's arguments on page 8 of the final Office Action that

it should be noted that Kubula discloses the speaker identification and segmentation system creating paragraph-like units between speakers and clustering archived files with a unique name. Thus, the Kubula's teaching of processing a clustered archived files with unique names is valid to read on the broadly claimed limitation of "identifying topics associated with the documents in the clusters,"

Appellants respectfully disagree. Speaker identification and segmentation allows the system to detect changes between speakers, which is important for correct playback of audio sections of an archive (Colbath et al., column 5, lines 19-23). Speaker identification has nothing to do with “means for identifying topics associated with the documents in the clusters,” as recited in claim 9.

Even assuming, for the sake of argument, that speaker identification and segmentation could reasonably be equated to identifying topics associated with the documents in the clusters (a point that Appellants do not concede), nowhere does Colbath et al. disclose or suggest that the “unique name” assigned to the archived files are formed from cluster lists that include topics that are associated with at least half of the documents in the clusters, as required by claim 9.

The Examiner continues to argue that Colbath et al. discloses the recited features of claim 9, but merely restates the previous rejection without explaining how the cited sections of Colbath et al. disclose the recited features of claim 9.

For at least the foregoing reasons, Appellants submit that the rejection of claim 9 under 35 U.S.C. § 102(b) based on Colbath et al. is improper. Accordingly, Appellants request that the rejection be reversed.

3. Claim 16

Claim 16 recites a topic detection system that includes a decision engine configured to receive a plurality of documents, and group the documents into a plurality of clusters; and a label engine configured to identify topics associated with the documents in the clusters, determine whether the topics are associated with at least half of the documents in the clusters, and form labels

for the clusters using ones of the topics that are associated with at least half of the documents in the clusters. Colbath et al. does not disclose or suggest this combination of features.

For example, Colbath et al. does not disclose clusters of documents. Colbath et al. mentions the word "cluster" in a few places in the context of clustering speakers together (see, e.g., column 5, lines 21-32), but does not disclose clusters of documents.

Because Colbath et al. does not disclose clusters of documents, Colbath et al. cannot disclose a label engine configured to determine whether topics are associated with at least half of the documents in the clusters, as recited in claim 16. The Examiner alleged that Colbath et al. discloses this feature and cited column 14, lines 1-26, of Colbath et al. for support (final Office Action, page 3). Appellants disagree.

At column 14, lines 1-26, Colbath et al. discloses:

One solution to this is to use traditional relevance feedback. The user has the option of specifying an entire story to the query system. When this is done, all the words in the story are fed into the full-text search engine, which returns five documents that use the maximum number of common terms with the seed document.

In the example in Fig. 7, we've given the system the first story in the "Smoking and FDA" query for a relevance feedback operation. The full-text search system has returned five stories. The first one is the seed story, since it has the most terms in common with itself. The second one happens to be the second-ranked story from the boolean query. The remaining three stories, however, are three stories on highly similar topics that weren't found with the boolean query mechanism. It should be emphasized that this mode of search becomes particularly important when the document source is text with errors introduced by a speech recognition system. Because of speech recognition errors, highly relevant documents may fall through the cracks of a boolean search, but are more likely to be found via relevance feedback since they will contain other words in common that are recognized correctly.

In this section, Colbath et al. discloses the use of traditional relevance feedback involving feeding all the words in a story into a full-text search engine, which returns five documents that use the maximum number of common terms with the seed document. Nowhere in this section, or

elsewhere, does Colbath et al. disclose or suggest a label engine configured to determine whether identified topics are associated with at least half of the documents in the clusters, as required by claim 1.

The Examiner also alleged that "three out of five search results contain highly similar topics for the cluster group" (final Office Action, page 3). Appellants cannot understand what the Examiner is alleging, but assert that, even if this allegation is reasonable, it has nothing to do with determining whether identified topics are associated with at least half of the documents in the clusters, as required by claim 1.

Colbath et al. also does not disclose a label engine configured to form labels for the clusters using ones of the topics that are associated with at least half of the documents in the clusters, as further recited in claim 16. Instead, Colbath et al. discloses forming a title for a document from all of the topics associated with that document (column 10, lines 13-20).

The Examiner alleged that Colbath et al. discloses a label engine configured to form labels for the clusters using ones of the topics that are associated with at least half of the documents in the clusters, and cited column 15, lines 27-37 of Colbath et al. for support (final Office Action, page 5). Appellants disagree.

At column 15, lines 27-37, Colbath et al. discloses:

There is no particular reason that this technique could not be extended to include proper name tagging, marking of new vocabulary words for the recognizer, or identification of new topics for the topic classifier. The latter is particularly important, since it is unlikely that an end-user could find a ready-made set of topics for their own meetings or teleconferences. For some particular problem domains, it may be sufficient to have a small set of topics (3-4 instead of the current 5,500).

In this section, Colbath et al. discloses implementing techniques to include proper name tagging, marking of new vocabulary words for a recognizer, or identification of new topics for a topic

classifier. Nowhere in this section, or elsewhere, does Colbath et al. disclose or suggest a label engine configured to form labels for the clusters using ones of the topics that are associated with at least half of the documents in the clusters, as required by claim 16.

For at least the foregoing reasons, Appellants submit that the rejection of claim 16 under 35 U.S.C. § 102(b) based on Colbath et al. is improper. Accordingly, Appellants request that the rejection be reversed.

B. Rejection under 35 U.S.C. § 103 based on Colbath et al. and Liddy et al.

The initial burden of establishing a *prima facie* basis to deny patentability to a claimed invention always rests upon the Examiner. In re Oetiker, 977 F.2d 1443, 24 USPQ2d 1443 (Fed. Cir. 1992). In rejecting a claim under 35 U.S.C. § 103, the Examiner must provide a factual basis to support the conclusion of obviousness. In re Warner, 379 F.2d 1011, 154 USPQ 173 (CCPA 1967). Based upon the objective evidence of record, the Examiner is required to make the factual inquiries mandated by Graham v. John Deere Co., 86 S.Ct. 684, 383 U.S. 1, 148 USPQ 459 (1966). The Examiner is also required to explain how and why one having ordinary skill in the art would have been realistically motivated to modify an applied reference and/or combine applied references to arrive at the claimed invention. Uniroyal, Inc. v. Rudkin-Wiley Corp., 837 F.2d 1044, 5 USPQ2d 1434 (Fed. Cir. 1988).

In establishing the requisite motivation, it has been consistently held that the requisite motivation to support the conclusion of obviousness is not an abstract concept, but must stem from the prior art as a whole to impel one having ordinary skill in the art to modify a reference or to combine references with a reasonable expectation of successfully achieving some particular realistic

objective. See, for example, Interconnect Planning Corp. v. Feil, 227 USPQ 543 (Fed. Cir. 1985). Consistent legal precedent admonishes against the indiscriminate combination of prior art references. Carella v. Starlight Archery, 804 F.2d 135, 231 USPQ 644 (Fed. Cir. 1986); Ashland Oil, Inc. v. Delta Resins & Refractories, Inc., 776 F.2d 281, 227 USPQ 657 (Fed. Cir. 1985).

1. Claim 4

Claim 4 depends from claim 1. Appellants submit that the disclosure of Liddy et al. does not remedy the deficiencies in the disclosure of Colbath et al. set forth with respect to claim 1. Therefore, Appellants request that the rejection of claim 4 be reversed for at least the reasons given above with respect to claim 1. Moreover, claim 4 recites additional features not disclosed or suggested by Colbath et al. and Liddy et al.

For example, claim 4 recites assigning ranks to the ones of the topics based on a number of the documents with which the ones of the topics are associated. Neither Colbath et al. nor Liddy et al. discloses the combination of features of claim 4.

The Examiner admitted that Colbath et al. does not disclose assigning ranks, but alleged that Liddy et al. discloses assigning ranks and cited column 21, lines 28-52, of Liddy et al. for support (final Office Action, page 5). Appellants disagree.

At column 21, lines 28-52, Liddy et al. discloses:

Matcher 55 matches documents by comparing the documents with the query and assigning each document a similarity score for the particular query. Documents with sufficiently high scores are arranged in ranked order in three folders, according to their relative relevance to the substance of a query. There are a number of evidence sources used for determining the similarity of documents to a query request, including:

Complex Nominals (CNs)*

Proper Nouns (PNs)*

Subject Field Codes (SFCs)

Single Terms*

Text Structure

Presence of Negation

Mandatory requirements

*CNs, PNs, and Single Terms are collectively called "terms."

Documents are arranged for the user based on a two-tier ranking system. The highest-level ranking mechanism is a system of folders. Documents are placed within folders based on various criteria, such as the presence or absence of mandatory terms. The lower-level ranking mechanism sorts documents within each folder based on criteria such as similarity score, document date assignment, etc.

In this section, Liddy et al. discloses ranking search result documents based on a query match.

Nowhere in this section, or elsewhere, does Liddy et al. disclose or suggest assigning ranks to topics, let alone assigning ranks to the ones of the topics based on a number of the documents with which the ones of the topics are associated, as required by claim 4.

On page 9 of the final Office Action, the Examiner argues that, “[a]lthough Kubula does not explicitly disclose all the claimed limitations, the feature not disclosed by Kubula is disclosed by Liddy. One can not show non-obviousness by attacking references individually where, as here, the rejection is based on a combination of references.” The Examiner, however, fails to explain how the cited sections of Colbath et al. and Liddy et al. disclose the recited features of claim 4.

For at least this additional reason, Appellants submit that the rejection of claim 4 under 35 U.S.C. § 103(a) based on Colbath et al. and Liddy et al. is improper. Accordingly, Appellants request that the rejection be reversed.

2. Claims 5-8

Claim 5 depends from claim 1. Appellants submit that the disclosure of Liddy et al. does not remedy the deficiencies in the disclosure of Colbath et al. set forth with respect to claim 1.

Therefore, Appellants request that the rejection of claim 5 be reversed for at least the reasons given above with respect to claim 1. Moreover, claim 5 recites additional features not disclosed or suggested by Colbath et al. and Liddy et al., whether taken alone or in any reasonable combination.

For example, claim 5 recites ranking the ones of the topics based on a number of the documents with which the ones of the topics are associated. Neither Colbath et al. nor Liddy et al. discloses the combination of features of claim 5.

The Examiner alleged that both Colbath et al. and Liddy et al. disclose these features and cited column 7, lines 7-15 and column 14, lines 8-26, of Colbath et al., and column 24, line 56 - column 25, line 2, of Liddy et al. for support (final Office Action, page 6). Appellants disagree.

At column 7, lines 7-15, Colbath et al. discloses:

Topic samples are taken from a sliding 200-word window across the transcribed text. Runs of similar high-ranking topics are combined to create *story* boundaries that give the user a high-level view of the data being shown, as well as providing a document model for information retrieval. The current set of approximately 5,500 topics come from an outside vendor, and apply specifically to broadcast news.

In this section, Colbath et al. discloses that runs of similar high-ranking topics are combined to create *story* boundaries that give the user a high-level view of the data being shown, as well as providing a document model for information retrieval. Nowhere in this section does Colbath et al. even mention ranking topics, let alone ranking the ones of the topics based on a number of the documents with which the ones of the topics are associated, as required by claim 5.

At column 14, lines 8-26, Colbath et al. discloses:

In the example in Fig. 7, we've given the system the first story in the "Smoking and FDA" query for a relevance feedback operation. The full-text search system has returned five stories. The first one is the seed story, since it has the most terms in common with itself. The second one happens to be the second-ranked story from the boolean query. The remaining three stories, however, are three stories on highly similar topics that weren't found with the boolean query mechanism. It should be emphasized that this mode of search becomes particularly important when the document source is text with errors introduced by a speech recognition system. Because of speech recognition errors, highly relevant documents may fall through the cracks of a boolean search, but are more likely to be found via relevance feedback since they will contain other words in common that are recognized correctly.

This section of Colbath et al. discloses the results of a full-text search system when the system is given a query for a relevance feedback operation. Nowhere in this section does Colbath et al. even mention ranking topics, let alone ranking the ones of the topics based on a number of the documents with which the ones of the topics are associated, as required by claim 5.

At column 24, line 56 - column 25, line 2, Liddy et al. discloses:

The matching of documents to a query organizes documents by matching scores in a ranked list. The total number of presented documents can be selected by the user, the system can determine a number using the Recall Predictor (RP) function, or, in the absence of user input, the system will retrieve all documents with a non-zero score. Note that documents from different sources are interfiled and ranked in a single list.

The RP filtering function is accomplished by means of a multiple regression formula that successfully predicts cut-off criteria on a ranked list of relevant documents for individual queries based on the similarity of documents to queries as indicated by the vector matching (and optionally the proper noun matching) scores.

In this section, Liddy et al. discloses matching documents to a search query and organizing the documents by matching scores in a ranked list. Nowhere in this section, or elsewhere, does Liddy et al. even mention ranking topics, let alone ranking the ones of the topics based on a number of the documents with which the ones of the topics are associated, as required by claim 5.

On page 9 of the final Office Action, the Examiner argues that, "[a]lthough Kubula does not explicitly disclose all the claimed limitations, the feature not disclosed by Kubula is disclosed by

Liddy. One can not show non-obviousness by attacking references individually where, as here, the rejection is based on a combination of references.” The Examiner, however, fails to explain how the cited sections of Colbath et al. and Liddy et al. disclose the recited features of claim 5.

For at least the foregoing reason, Appellants submit that the rejection of claim 5 under 35 U.S.C. § 103(a) based on Colbath et al. and Liddy et al. is improper. Accordingly, Appellants request that the rejection be reversed.

Claims 6-8 depend from claim 5. Therefore, Appellants request that the rejection of claims 6-8 be reversed for at least the reasons given above with respect to claim 5.

3. Claims 10 and 11

Claims 10 and 11 depend from claim 9. Appellants submit that the disclosure of Liddy et al. does not remedy the deficiencies in the disclosure of Colbath et al. set forth with respect to claim 9. Therefore, Appellants request that the rejection of claims 10 and 11 be reversed for at least the reasons given above with respect to claim 9.

4. Claim 17

Claim 17 depends from claim 16. Appellants submit that the disclosure of Liddy et al. does not remedy the deficiencies in the disclosure of Colbath et al. set forth with respect to claim 16. Therefore, Appellants request that the rejection of claim 17 be reversed for at least the reasons given above with respect to claim 16. Moreover, claim 17 recites additional features not disclosed or suggested by Colbath et al. and Liddy et al.

For example, claim 17 recites that the label engine is further configured to rank the ones of the topics based on a number of the documents with which the ones of the topics are associated.

Neither Colbath et al. nor Liddy et al. discloses the combination of features of claim 17.

The Examiner alleged that both Colbath et al. and Liddy et al. disclose these features and cited column 7, lines 7-15 and column 14, lines 8-26, of Colbath et al., and column 24, line 56 - column 25, line 2, of Liddy et al. for support (final Office Action, page 6). Appellants disagree.

At column 7, lines 7-15, Colbath et al. discloses:

Topic samples are taken from a sliding 200-word window across the transcribed text. Runs of similar high-ranking topics are combined to create *story* boundaries that give the user a high-level view of the data being shown, as well as providing a document model for information retrieval. The current set of approximately 5,500 topics come from an outside vendor, and apply specifically to broadcast news.

In this section, Colbath et al. discloses that runs of similar high-ranking topics are combined to create *story* boundaries that give the user a high-level view of the data being shown, as well as providing a document model for information retrieval. Nowhere in this section does Colbath et al. even mention ranking topics, let alone that the label engine is further configured to rank the ones of the topics based on a number of the documents with which the ones of the topics are associated, as required by claim 5.

At column 14, lines 8-26, Colbath et al. discloses:

In the example in Fig. 7, we've given the system the first story in the "Smoking and FDA" query for a relevance feedback operation. The full-text search system has returned five stories. The first one is the seed story, since it has the most terms in common with itself. The second one happens to be the second-ranked story from the boolean query. The remaining three stories, however, are three stories on highly similar topics that weren't found with the boolean query mechanism. It should be emphasized that this mode of search becomes particularly important when the document source is text with errors introduced by a speech recognition system. Because of speech recognition errors, highly relevant documents may fall through the cracks of a boolean search, but are more likely to be found via relevance feedback since they will contain other words in common that are recognized correctly.

This section of Colbath et al. discloses the results of a full-text search system when the system is given a query for a relevance feedback operation. Nowhere in this section does Colbath et al. even mention ranking topics, let alone that the label engine is further configured to rank the ones of the topics based on a number of the documents with which the ones of the topics are associated, as required by claim 17.

At column 24, line 56 - column 25, line 2, Liddy et al. discloses:

The matching of documents to a query organizes documents by matching scores in a ranked list. The total number of presented documents can be selected by the user, the system can determine a number using the Recall Predictor (RP) function, or, in the absence of user input, the system will retrieve all documents with a non-zero score. Note that documents from different sources are interfiled and ranked in a single list.

The RP filtering function is accomplished by means of a multiple regression formula that successfully predicts cut-off criteria on a ranked list of relevant documents for individual queries based on the similarity of documents to queries as indicated by the vector matching (and optionally the proper noun matching) scores.

In this section, Liddy et al. discloses matching documents to a search query and organizing the documents by matching scores in a ranked list. Nowhere in this section, or elsewhere, does Liddy et al. even mention ranking topics, let alone that the label engine is further configured to rank the ones of the topics based on a number of the documents with which the ones of the topics are associated, as required by claim 17.

On page 9 of the final Office Action, the Examiner argues that, “[a]lthough Kubula does not explicitly disclose all the claimed limitations, the feature not disclosed by Kubula is disclosed by Liddy. One can not show non-obviousness by attacking references individually where, as here, the rejection is based on a combination of references.” The Examiner, however, fails to explain how the cited sections of Colbath et al. and Liddy et al. disclose the recited features of claim 17.

For at least the foregoing reason, Appellants submit that the rejection of claim 17 under 35 U.S.C. § 103(a) based on Colbath et al. and Liddy et al. is improper. Accordingly, Appellants request that the rejection be reversed.

C. Rejection under 35 U.S.C. § 102 or 35 U.S.C. § 103 based on Colbath et al.

1. Claims 18 and 19

Independent claim 18 is directed to a method for creating labels for clusters of documents. The method comprises identifying topics associated with the documents in the clusters; determining whether the topics are associated with at least a predetermined number of the documents in the clusters; and generating labels for the clusters using ones of the topics that are associated with the at least a predetermined number of the documents in the clusters.

Colbath et al. does not disclose the combination of features recited in claim 18. For example, Colbath et al. does not disclose determining whether identified topics are associated with at least a predetermined number of the documents in the clusters. The Examiner alleged that Colbath et al. discloses this feature and cited column 14, lines 1-26 of Colbath et al. for support (final Office Action, page 7). Appellants disagree.

At column 14, lines 1-26, Colbath et al. discloses:

One solution to this is to use traditional relevance feedback. The user has the option of specifying an entire story to the query system. When this is done, all the words in the story are fed into the full-text search engine, which returns five documents that use the maximum number of common terms with the seed document.

In the example in Fig. 7, we've given the system the first story in the "Smoking and FDA" query for a relevance feedback operation. The full-text search system has returned five stories. The first one is the seed story, since it has the most terms in common with itself. The second one happens to be the second-ranked story from the boolean query. The remaining three stories, however, are three stories on highly similar topics that weren't found

with the boolean query mechanism. It should be emphasized that this mode of search becomes particularly important when the document source is text with errors introduced by a speech recognition system. Because of speech recognition errors, highly relevant documents may fall through the cracks of a boolean search, but are more likely to be found via relevance feedback since they will contain other words in common that are recognized correctly.

In this section, Colbath et al. discloses the use of traditional relevance feedback involving feeding all the words in a story into a full-text search engine, which returns five documents that use the maximum number of common terms with the seed document. Nowhere in this section, or elsewhere, does Colbath et al. disclose determining whether identified topics are associated with at least a predetermined number of the documents in the clusters, as required by claim 18.

The Examiner also alleged that "three out of five search results contain highly similar topics for the cluster group" (final Office Action, page 7). Appellants cannot understand what the Examiner is alleging, but assert that, even if this allegation is reasonable, it has nothing to do with determining whether identified topics are associated with at least half of the documents in the clusters, as required by claim 1.

Colbath et al. also does not disclose generating labels for the clusters using ones of the topics that are associated with the at least a predetermined number of the documents in the clusters, as further recited in claim 18. Instead, Colbath et al. discloses forming a title for a document from all of the topics associated with that document (column 10, lines 13-20).

The Examiner alleged that Colbath et al. discloses generating labels for the clusters using ones of the topics that are associated with the at least a predetermined number of the documents in the clusters, and cited column 8, lines 11-16, and column 14, line 8 – column 15, line 37, of Colbath et al. for support (final Office Action, page 7). Appellants disagree.

At column 8, lines 11-16, Colbath et al. discloses:

The Rough'n'Ready IR system uses a full-text search system developed at BBN which uses an HMM-based model of document retrieval. This system, described in [7], is used in relevance-feedback mode to allow the user of the system to find documents that are similar to an exemplar.

In this section, Colbath et al. discloses an HMM-based model for document retrieval. Nowhere in this section, or elsewhere, does Colbath et al. disclose generating labels for the clusters using ones of the topics that are associated with the at least a predetermined number of the documents in the clusters, as required by claim 18.

At column 14, lines 29 - column 15, line 37, Colbath et al. discloses:

Annotation

There is no particular reason that the database has to be browsed in a read-only fashion, however. The training data for the Rough'n'Ready indexer is currently fairly static. To annotate more speech data, or additional names for the name spotter, or additional topics for the topic classifier is a separate, offline process using dedicated annotators. However, since the current annotation process is relatively simple and does not require any in-depth linguistic knowledge, it seems logical that the enduser of the archive should be enlisted in helping to provide the training data. This makes sense since it is likely the consumer of the data will have the most familiar with it, and will be able to provide topics, identify speakers, etc.

The current Rough'n'Ready system includes some basic speaker annotation capabilities. If the user encounters a speaker currently marked as unknown, they can step through a relatively simple wizard that will play segments of data that have been tagged with the same identifier (such as "Male 5") and ask them to confirm that this is the same as the first speaker. Once they have accumulated enough data (three to five minutes), the system trains a new speaker model, and reprocesses the rest of the archive off-line to include the new speaker. It is also possible to add extra training data for speakers that have particularly weak performance, improving their models.

There is no particular reason that this technique could not be extended to include proper name tagging, marking of new vocabulary words for the recognizer, or identification of new topics for the topic classifier. The latter is particularly important, since it is unlikely that an end-user could find a ready-made set of topics for their own meetings or teleconferences. For some particular problem domains, it may be sufficient to have a small set of topics (3-4 instead of the current 5,500).

In this section, Colbath et al. discloses that annotators can annotate more speech data, additional names for the name spotter, or additional topics for the topic classifier. This section of Colbath et al. also discloses including proper name tagging, marking of new vocabulary words for a recognizer, and identification of new topics for a topic classifier. Nowhere in this section, or elsewhere, does Colbath et al. disclose generating labels for the clusters using ones of the topics that are associated with the at least a predetermined number of the documents in the clusters, as required by claim 18.

The Examiner also alleged that Colbath et al. discloses a "new identification of new topics for the topic classifier" (final Office Action, page 7). Even assuming, for the sake of argument, that the Examiner's allegation is reasonable (a point that Appellants do not concede), the Examiner has not addressed the feature of generating labels for the clusters using ones of the topics that are associated with the at least a predetermined number of the documents in the clusters, as required by claim 18.

For at least the foregoing reasons, Appellants submit that the rejection of claim 18 under 35 U.S.C. § 102(b) or 35 U.S.C. § 103(a) based on Colbath et al. is improper. Accordingly, Appellants request that the rejection be reversed.

Claim 19 depends from claim 18. Therefore, Appellants request that the rejection of claim 19 be reversed for at least the reasons given above with respect to claim 18.

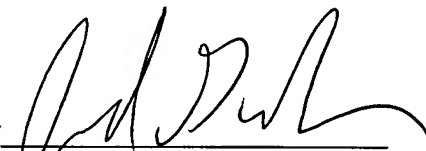
VIII. CONCLUSION

In view of the foregoing arguments, Appellant respectfully solicits the Honorable Board to reverse the Examiner's rejections of claims 1-11 and 16-19.

Applicant believes no fee is due with this response. However, if a fee is due, please charge our Deposit Account No. 18-1945, under Order No. BBNT-P01-199 from which the undersigned is authorized to draw.

Dated: August 6, 2007

Respectfully submitted,

By 

Edward A. Gordon

Registration No.: 54,130

FISH & NEAVE IP GROUP, ROPES & GRAY
LLP

1211 Avenue of the Americas

New York, New York 10036

(212) 596-9000

(212) 596-9090 (Fax)

Attorneys/Agents For Applicant

IX. APPENDIX

1. A method of creating labels for clusters of documents, comprising:
identifying topics associated with the documents in the clusters;
determining whether the topics are associated with at least half of the documents in the clusters;
adding ones of the topics that are associated with at least half of the documents in the clusters to cluster lists; and
forming labels for the clusters from the cluster lists.
2. The method of claim 1, wherein the identifying topics includes:
using a probabilistic Hidden Markov Model to determine the topics.
3. The method of claim 1, wherein the forming labels includes:
ranking the ones of the topics, and
placing the ones of the topics in the labels in ranked order.
4. The method of claim 3, wherein the ranking the ones of the topics includes:
assigning ranks to the ones of the topics based on a number of the documents with which the ones of the topics are associated.
5. The method of claim 1, further comprising:

ranking the ones of the topics based on a number of the documents with which the ones of the topics are associated.

6. The method of claim 5, wherein when a first one of the ones of the topics, as a first topic, is associated with a majority of the documents in one of the clusters and a second one of the ones of the topics, as a second topic, is associated with less than the majority of the documents in the one of the clusters, the first topic is ranked higher than the second topic.

7. The method of claim 5, wherein the ranking the ones of the topics includes:
assigning higher ranks to first ones of the ones of the topics that are associated with larger numbers of the documents than second ones of the ones of the topics that are associated with smaller numbers of the documents.

8. The method of claim 5, wherein the forming labels includes:
sorting the cluster lists based on the rankings of the ones of the topics.

9. A system for generating a label for a cluster of documents, comprising:
means for identifying topics associated with the documents in the cluster;
means for determining whether the topics are associated with at least half of the documents in the cluster; and
means for generating a label for the cluster based on one or more of the topics that are associated with at least half of the documents in the cluster.

10. The system of claim 9, further comprising:
means for ranking the one or more of the topics based on a number of the documents with which the one or more of the topics are associated.

11. The system of claim 10, wherein the means for generating a label includes:
means for sorting the one or more of the topics based on the ranking to form the label for the cluster.

12. A system for creating a label for a cluster of documents, comprising:
logic configured to identify topics associated with the documents in the cluster;
logic configured to determine whether the topics are associated with approximately half or more of the documents in the cluster;
logic configured to rank ones of the topics that are associated with approximately half or more of the documents in the cluster; and
logic configured to generate a label for the cluster using the ones of the topics in ranked order.

13. The system of claim 12, wherein when a first one of the ones of the topics, as a first topic, is associated with a majority of the documents in the cluster and a second one of the ones of the topics, as a second topic, is associated with less than the majority of the documents in the cluster, the first topic is ranked higher than the second topic.

14. The system of claim 12, wherein the logic configured to rank ones of the topics includes:

logic configured to assign higher ranks to first ones of the ones of the topics that are associated with larger numbers of the documents than second ones of the ones of the topics that are associated with smaller numbers of the documents.

15. The system of claim 12, wherein the logic configured to generate a label includes: logic configured to sort the ones of the topics based on the rankings of the ones of the topics.

16. A topic detection system, comprising:

a decision engine configured to:

receive a plurality of documents, and

group the documents into a plurality of clusters; and

a label engine configured to:

identify topics associated with the documents in the clusters,

determine whether the topics are associated with at least half of the documents in the clusters, and

form labels for the clusters using ones of the topics that are associated with at least half of the documents in the clusters.

17. The system of claim 16, wherein the label engine is further configured to:

rank the ones of the topics based on a number of the documents with which the ones of the topics are associated.

18. A method for creating labels for clusters of documents, comprising:
identifying topics associated with the documents in the clusters;
determining whether the topics are associated with at least a predetermined number of the documents in the clusters; and
generating labels for the clusters using ones of the topics that are associated with the at least a predetermined number of the documents in the clusters.

19. The method of claim 18, wherein the predetermined number of the documents is equal to approximately half of the documents.

X. EVIDENCE APPENDIX

None

XI. RELATED PROCEEDINGS APPENDIX

None

COPY



Via: First Class Mail	Atty Dkt No.: BBNT-P01-199
Inventor: Colbath et al.	
Application No.: 10/685479	Filing Date: October 16, 2003
Title: SYSTEMS AND METHODS FOR LABELING CLUSTERS OF DOCUMENTS	

Documents Filed:
Transmittal (1 page)

Revocation of Power of Attorney With New Power of Attorney and Change of
Correspondence Address (1 page)
Statement Under 37 CFR 3.73(b) (1 page)
Change of Attorney Docket Number (1 page)
Deposit Account 18-1945
Return receipt postcard

Sender's Initials: EAG/r Date: April 28, 2005

Via: First Class Mail	Atty Dkt No.: BBNT-P01-199
Inventor: Colbath et al.	
Application No.: 10/685479	Filing Date: October 16, 2003
Title: SYSTEMS AND METHODS FOR LABELING CLUSTERS OF DOCUMENTS	

Documents Filed:
Transmittal (1 page)

Revocation of Power of Attorney With New Power of Attorney and Change of
Correspondence Address (1 page)
Statement Under 37 CFR 3.73(b) (1 page)

Change of Attorney Docket Number (1 page)

Deposit Account 18-1945

Return receipt postcard

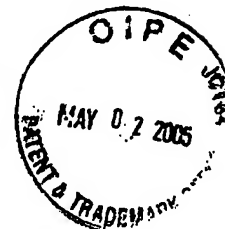
Ropes & Gray

MAY 09 2005

Sender's Initials: EAG/r

Date: April 28, 2005

Intellectual Property Dept.



COPY



PTO/SB/21 (09-04)

Approved for use through 07/31/2006. OMB 0651-0031

U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

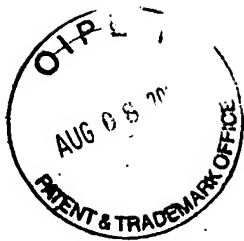
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

<h2>TRANSMITTAL FORM</h2> <p>(to be used for all correspondence after initial filing)</p>		Application Number	10/685479
		Filing Date	October 16, 2003
		First Named Inventor	Sean Colbath
		Art Unit	2176
		Examiner Name	Not Yet Assigned
Total Number of Pages in This Submission	5	Attorney Docket Number	BBNT-P01-199

ENCLOSURES (Check all that apply)		
<input type="checkbox"/> Fee Transmittal Form <input type="checkbox"/> Fee Attached <input type="checkbox"/> Amendment/Reply <input type="checkbox"/> After Final <input type="checkbox"/> Affidavits/declaration(s) <input type="checkbox"/> Extension of Time Request <input type="checkbox"/> Express Abandonment Request <input type="checkbox"/> Information Disclosure Statement <input type="checkbox"/> Certified Copy of Priority Document(s) <input type="checkbox"/> Reply to Missing Parts/Incomplete Application <input type="checkbox"/> Reply to Missing Parts under 37 CFR 1.52 or 1.53	<input type="checkbox"/> Drawing(s) <input type="checkbox"/> Licensing-related Papers <input type="checkbox"/> Petition <input type="checkbox"/> Petition to Convert to a Provisional Application <input type="checkbox"/> Power of Attorney, Revocation Change of Correspondence Address <input type="checkbox"/> Terminal Disclaimer <input type="checkbox"/> Request for Refund <input type="checkbox"/> CD, Number of CD(s) _____ <input type="checkbox"/> Landscape Table on CD	<input type="checkbox"/> After Allowance Communication to TC <input type="checkbox"/> Appeal Communication to Board of Appeals and Interferences <input type="checkbox"/> Appeal Communication to TC (Appeal Notice, Brief, Reply Brief) <input type="checkbox"/> Proprietary Information <input type="checkbox"/> Status Letter <input checked="" type="checkbox"/> Other Enclosure(s) (please identify below): Revocation of Power of Attorney With New Power of Attorney and Change of Correspondence Address Statement Under 3.73(b) Change of Attorney Docket Number Return Receipt Postcard
<div>Remarks</div>		

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT			
Firm Name	ROPES & GRAY LLP		
Signature			
Printed name	Edward A. Gordon		
Date	April 28, 2005	Reg. No.	54,130

I hereby certify that this correspondence is being deposited with the U.S. Postal Service with sufficient postage as First Class Mail, in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on the date shown below.	
Dated: 4/28/05	Signature: (Joanne Ryan)



COPY

PTO/SB/82 (09-04)
Approved for use through 11/30/2005. OMB 0651-0036
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

REVOCATION OF POWER OF ATTORNEY WITH NEW POWER OF ATTORNEY AND CHANGE OF CORRESPONDENCE ADDRESS	Application Number	10/685479
	Filing Date	October 16, 2003
	First Named Inventor	Sean Colbath
	Art Unit	2176
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	BBNT-P01-199

I hereby revoke all previous powers of attorney given in the above-identified application.

☐ A Power of Attorney is submitted herewith.
OR
☒ I hereby appoint the practitioners associated with the Customer Number: 28120

☒ Please change the correspondence address for the above-identified application to:
☒ The address associated with Customer Number: 28120
OR

<input type="checkbox"/> Firm or Individual Name	ROPES & GRAY LLP Edward J. Kelly		
Address	One International Place		
City	Boston		
Country	US	State	MA
		Zip	02110-2624
Telephone	(617) 951-7000	Fax	(617) 951-7050

I am the:

☐ Applicant/Inventor.
☒ Assignee of record of the entire interest. See 37 CFR 3.71.
Statement under 37 CFR 3.73(b) is enclosed. (Form PTO/SB/96)

SIGNATURE of Applicant or Assignee of Record

Signature			
Name	Mark Sherman		
Date	4/21/05	Telephone	

NOTE: Signatures of all the inventors or assignees of record of the entire interest or their representative(s) are required. Submit multiple forms if more than one signature is required, see below*.

☐ *Total of 1 forms are submitted.

I hereby certify that this correspondence is being deposited with the U.S. Postal Service with sufficient postage as First Class Mail, in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on the date shown below.	
Dated: <u>4/28/05</u>	Signature: <u>Joanne Ryan</u> (Joanne Ryan)

COPY



Approved for use through 07/31/2008. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT UNDER 37 CFR 3.73(b)

Applicant/Patent Owner: Colbath et al.

Application No./Patent No.: 10/685479

Filed/Issue Date: October 16, 2003

Entitled: SYSTEMS AND METHODS FOR LABELING CLUSTERS OF DOCUMENTS

BBNT Solutions LLC
(Name of Assignee)

, a Limited liability company
(Type of Assignee, e.g., corporation, partnership, university, government agency, etc.)

states that it is:

1. ☒ the assignee of the entire right, title, and interest; or
2. ☐ an assignee of less than the entire right, title and interest.
The extent (by percentage) of its ownership interest is _____ %

In the patent application/patent identified above by virtue of either:

- A. ☒ An assignment from the inventor(s) of the patent application/patent identified above. The assignment was recorded in the United States Patent and Trademark Office at Reel 015815, Frame 0330, or for which a copy thereof is attached.

OR

- B. ☐ A chain of title from the inventor(s), of the patent application/patent identified above, to the current assignee as shown below:

1. From: _____ To: _____
The document was recorded in the United States Patent and Trademark Office at Reel _____, Frame _____, or for which a copy thereof is attached.
2. From: _____ To: _____
The document was recorded in the United States Patent and Trademark Office at Reel _____, Frame _____, or for which a copy thereof is attached.
3. From: _____ To: _____
The document was recorded in the United States Patent and Trademark Office at Reel _____, Frame _____, or for which a copy thereof is attached.

☐ Additional documents in the chain of title are listed on a supplemental sheet.

☐ Copies of assignments or other documents in the chain of title are attached.
[NOTE: A separate copy (i.e., a true copy of the original assignment document(s)) must be submitted to Assignment Division in accordance with 37 CFR Part 3, if the assignment is to be recorded in the records of the USPTO. See MPEP 302.08]

The undersigned (whose title is supplied below) is authorized to act on behalf of the assignee.

Mark Sherman
Signature
Mark Sherman
Printed or Typed Name
Authorized Signer for Assignee
Title

4/21/05
Date
Telephone Number

I hereby certify that this correspondence is being deposited with the U.S. Postal Service with sufficient postage as First Class Mail, in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on the date shown below.

Dated: 4/28/05 Signature: Joanne Ryan (Joanne Ryan)



COPY

I hereby certify that this correspondence is being deposited with the U.S. Postal Service with sufficient postage as First Class Mail, in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on the date shown below.

Dated: 4/28/05

Signature: Joanne Ryan
(Joanne Ryan)

Docket No.: BBNT-P01-199
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:
Colbath et al.

Application No.: 10/685479

Group Art Unit: 2176

Filed: October 16, 2003

Examiner: Not Yet Assigned

For: SYSTEMS AND METHODS FOR LABELING
CLUSTERS OF DOCUMENTS

CHANGE OF ATTORNEY DOCKET NUMBER

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

Please note that the Attorney Docket Number has been changed from 02-4033 to **BBNT-P01-199**. Please reference **BBNT-P01-199** on all future correspondence.

Applicant believes no fee is due with this response. However, if a fee is due, please charge our Deposit Account No. 18-1945, under Order No. BBNT-P01-199 from which the undersigned is authorized to draw.

Dated: April 28, 2005

Respectfully submitted,

By Edward A. Gordon

Edward A. Gordon

Registration No.: 54,130

ROPES & GRAY LLP

One International Place

Boston, Massachusetts 02110-2624

(617) 951-7000

(617) 951-7050 (Fax)

Attorneys/Agents For Applicant